A STUDY OF CHI-SQUARE AND KOLMOGOROV-SMIRNOV TESTS

by

CHONG JIN PARK

B.S.,B.A., University of Washington, 1961, 1962

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics and
Statistical Laboratory

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1963

Approved by

H. C. Fryer

Major Professor

CONTENTS

INTRODUCTION

In standard applications of tests for goodness of fit, the chi-square test (denoted cst hereafter) and the Kolmogorov-Sminov test (denoted kst hereafter) are widely used. The cst can be applied in situations where the population has either a continuous or discrete distribution. On the other hand, the kst can be correctly used only in situation where the population has a continuous distribution.

Since the kst is based on the assumption of a continuous distribution, it is necessary to study whether this test may be applied in a situation in which the distribution is discrete.

For small samples from a hypothetical binomial population, the cst statistic is compared with the kst statistic. The comparision between the two tests was extended to large samples from hypothetical multinomial population.

As the probability distribution function is completely specified in this study, estimation of parameters was not considered.

CHI-SQUARE TEST FOR GOODNESS OF FIT

1.1. Chi-square Test Statistic and Its Asymptotic Distribution

The n observations $(x_1, x_2, \ldots, x_n)$ in a random sample from a population are classified into k+1 mutually exclusive classes. There is some theoretical probability function which specifies the probability p_i that an observation falls into the ith class. Sometimes they are completely specified by the probability function, sometimes they are less completely specified.

If the theoretical probability function is correct, observed numbers follow a multinomial distribution with p_i as the probability in the ith class. The joint distribution of the observations is therefore specified by the probability function;

$$p\,(x_1, x_2, \ldots, x_k) = \frac{n!}{x_1!\, x_2! \ldots x_{k+1}!}\, p_1^{x_1}\, p_2^{x_2} \ldots p_{k+1}^{x_{k+1}} \quad (1.1)$$

where $x_{k+1} = n - x_1 - x_2 \ldots - x_k$, and $p_{k+1} = 1 - p_1 - p_2 \ldots - p_k$.

One wants to test the null hypothesis that the observations are a random sample from the population with specified probability distribution. As a test criterion for the null hypothesis, Karl Pearson (1) proposed the following test statistic:

$$X^2 = \sum_{i=1}^{k+1} \frac{(x_i - np_i)^2}{np_i} \quad (1.2)$$

X^2 is a quadratic form in random variables $(x_i - np_i)$, i = 1, 2, ..., k+1, with the coefficient matrix being the inverse of the covariance matrix of multinomial distribution. Therefore another expression of X^2 is (2);

$$X^2 = \sum_{i,j=1}^{k+1} \sigma^{ij} \frac{(x_i - np_i)}{n} \frac{(x_j - np_j)}{n} \tag{1.2a}$$

where

$$(\sigma^{ij}) = (p_i \delta_{ij} - p_i p_j)^{-1} = \left(\frac{\delta_{ij}}{p_i} + \frac{1}{p_{k+1}} \right)$$

where

$\delta_{ij} = 1$ if $i = j$,

$\delta_{ij} = 0$ if $i \neq j$.

Hence if the null hypothesis is true, the limiting distribution of (1.2), as $n \to \infty$, is the chi-square distribution, with k degrees of freedom, whose probability density function is (3);

$$f(u) = \frac{\left(\frac{u}{2}\right)^{\frac{k}{2}-1}}{\left(\frac{k}{2}\right)} e^{-\frac{u}{2}} \quad , \qquad u > 0 \tag{1.3}$$

In practice, however, X^2 given in (1.2) is computed on the basis of random sample; and for large n, X^2 is assumed to have chi-square distribution, hence one uses its table (4) to obtain the probability;

$$P(X^2 \geqslant c) = \int_c^\infty f(u)du \tag{1.4}$$

where f(u) is the probability density function given in (1.3).

1.2. Chi-square Test Statistic for the Binomial Distribution

If a random sample of size n is drawn from the binomial distribution B(1;p), then the sample sum x has the binomial distribution B(n;p). Hence one obtains a test statistic from (1.2) for this sample sum x, and it can be written as;

$$X^2b = \frac{(x-np)^2}{np} + \frac{(n-x-n(1-p))^2}{n(1-p)} = \frac{(x-np)^2}{np(1-p)} \qquad (1.5)$$

Making use of the Table of the Binomial Distribution, one obtains the cumulative distribution of X^2b, namely for a given c one can find k such that;

$$P(X^2b \geqslant c) = P(x \leqslant k) = \sum_{y=0}^{k} \binom{n}{y} p^y (1-p)^{n-y} \qquad (1.6)$$

The cumulative distribution of X^2b is tabulated for n = 5, 10, 15, 20, 25, 30, and p=1/2, in TABLE I on page 12.

Figure 1. Comparision of X_b^2 and chi-square(X^2)distributions with one degree of freedom

n=20		n=30		chi-square	
c	$P(X^2b \geqslant c)$	c	$P(X^2b \geqslant c)$	c	$P(X^2 \geqslant c)$
1.800	.2632	2.133	.2004	2.706	.10
3.200	.1154	3.333	.0988	3.841	.05
5.000	.0414	4.800	.0428	5.412	.02
7.200	.0118	6.500	.0162	6.635	.01
9.800	.0026	8.533	.0052	10.827	.001
		10.800	.0014		

If the null hypothesis is true, as stated in section 1.1., X^2b has as its limiting distribution the chi-square distribution with 1 degree of freedom.

From TABLE I and Figure 1, as n increases, one notes that the exact distribution of X^2b is a fairly good approximation to the chi-square distribution with 1 degree of freedom. In other words, the exact probabilities associated with X^2b statistics, for large n, are good approximation of the probabilities associated with the random variable whose probability density function is given in (1.3) with 1 degree of freedom.

1.3. Chi-square Test Statistic for the Multinomial Distribution

If a random sample of size n is drawn from the multinomial distribution $M(1;p_1, p_2, \ldots, p_k)$, then the sample sum $(x_1, x_2, \ldots, x_k)$ has the multinomial distribution $M(n;p_1, p_2, \ldots, p_k)$, whose probability function is given in (1.1). The probability p_i that an observation falls into ith class is defined as follows;

$$p_k = \binom{10}{k-1} \left(\tfrac{1}{2}\right)^{10} \quad , \quad k = 1, 2, \ldots, 10. \qquad (1.7)$$

Hence the test statistic (1.2) for this sample can be written as;

$$X^2m = \sum_{i=1}^{11} \frac{(x_i - np_i)^2}{np_i} \qquad (1.8)$$

where $p_{11} = 1 - p_1 - p_2 \ldots - p_{10}$.

The cumulative distribution of X^2m may be obtained by the use of the probability function of $(x_1, x_2, \ldots, x_{10})$, but it will be very cumbersome since the number of classes is so large. Hence the Monte Carlo technique (5) was applied to get the approximate distribution of X^2m.

Two examples were considered; one for sample size 1024 and other for sample size 512. The computer (IBM 1620) was used to generate the hypothetical multinomial distribution $M(1;p_1, p_2, \ldots, p_{10})$ with p_i's being specified in (1.6)[1], hence the sample sum $(x_1, x_2, \ldots, x_{10})$ has $M(n;p_1, p_2, \ldots, p_{10})$ for n=1024 and n=512. This sampling and computation of X^2m were repeated a hundred times.

If the probabilities defined in (1.7) are true, the expected number of observations in each class will be as follows;

---

[1] See Appendix

n = 1024:	1	10	45	120	210	252	210	120	45	10	1
n = 512:	.5	5	22.5	60	105	126	105	60	22.5	5	.5

Since the expected numbers in the classes of extreme ends are too small for good approximation they were grouped with the adjacent ones (3).

If the null hypothesis is true, that is if the sample sum $(x_1, x_2, \ldots, x_{10})$ has multinomial distribution with p_i's being given by (1.7), the limiting distribution of (1.8), is the chi-square distribution with 8 degrees of freedom.

The sample cumulative distribution of X^2m was tabulated in TABLE II on page 15, for both n's. From this table it is clear that the distribution of X^2m is close to the chi-square distribution with 8 degrees of freedom. Another interpretation of this result is to say that the hypothetical distribution so generated is the specified multinomial distribution.

KOLMOGOROV-SMIRNOV TEST FOR GOODNESS OF FIT

2.1. Kolmogorov-Smirnov Test Statistic and Its Asymptotic Distribution

Let $(x_1, x_2, \ldots, x_n)$ be the n observations in a random sample from a population with a continuous cumulative distribution function F(x), which is completely specified. Define $Sn = N(x)/n$, where N(x) denotes the number of x_i's whose observed values do not exceed x.

Since F(x) is assumed to be continuous, Sn(x) is a step function with the magnitude of jumps at each x being 1/n. When n is large it is certain that Sn(x) of the sample will be approximately equal to the F(x). As the test criterion for null hypothesis that the sample is drawn from the population with cumulative distribution function F(x), A. Kolmogorov (6) proposed the test statistic;

$$Dn = \sup_{-\infty < x < \infty} | Sn(x) - F(x) | \qquad (2.1)$$

where sup is the abbreviation for supremum.

If F(x) is continuous, this test statistic has the great advantage that its distribution is independent of F(x). For this reason, the kst is a distribution-free statistic.

Let $y = F(x)$ and $yn = Sn(x)$. Then, because F(x) is continuous, y has the rectangular distribution R(1/2, 1); and the cumulative sample distribution Gn(y) of yn is a step function with n jumps of magnitude 1/n at each y. From these facts, the cumulative distribution function of the test statistics Dn can be written as;

$$Kn(k/\sqrt{n}) = P(Dn \leqslant k/\sqrt{n}) = P\left(\sup_{0<y<1} |Gn(y) - G(y)| \leqslant k/\sqrt{n} \right) \qquad (2.2)$$

Let $I_1, I_2, \ldots\ldots, I_n$ be n intervals defined on (0, 1] as $I_i = (x-1)/n, x/n]$ where $x = 1, 2, \ldots, n$. Let $(r_1, r_2, \ldots, r_n)$ be a random variable (degenerated with $r_1 + r_2 + \ldots\ldots + r_n = n$) denoting the numbers in the sample $y_1, y_2, \ldots\ldots, y_n$ falling into $I_1, I_2, \ldots, I_n$, respectively. The r's, of course, have an n-1 dimensional multinomial distribution whose probability function is given by

$$p(r_1, r_2, \ldots, r_n) = \frac{n!}{r_1!\, r_2! \ldots r_n!} \left(\frac{1}{n}\right)^n \qquad (2.3)$$

Now the random variable $(r_1, r_2, \ldots, r_n)$ uniquely determines Gn(y), hence the value of $P(Dn \leqslant k/\sqrt{n})$ is determined accordingly by summing (2.3) over all points in the sample space of $(r_1, r_2, \ldots, r_n)$ for which $|Gn(y) - G(y)| \leqslant k/\sqrt{n}$ for all y.

When n is large the distribution function given in (2.2) tends to

$$K(h) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-kh^2} \qquad (2.4)$$

uniformly with respect to h (10). Some of distributions of (2.4) has been tabulated by Smirnov (11). The distribution of this statistic for finite n given in (2.2) has been tabulated by Massey (7), and Birnbaum and Tingey (8), (9).

Without the assumption that F(x) is continuous, however $Kn(k/\sqrt{n})$ in (2.2) has its limiting distribution K(h) in (2.4) (6). But the limiting distributions are no longer independent of F(x). They depend on the value of F(x) at the discontinuity points; but not on the form of the function between the points of discontinuity (12).

2.2. Kolmogorov-Smirnov Test Statistic for the Binomial Distribution

As mention in section 2.1., the distribution of Dn is based on the assumption of a continuous F(x). It was, however, of interest to see how good the kst was if one applied it to a discrete distribution.

Consider the sample sum x of a random sample from the binomial distribution B(1;p) described in section 1.2.. From (2.1), the test statistic for this sample can be written as;

$$Db = \sup_x |x/n - p| \qquad (2.5)$$

where $p = \frac{1}{2}$.

Making use of the Table of the Binomial Distribution together with (2.5), the cumulative distribution of Db for n=5, 10, 15, 20, 25, 30, and $p = \frac{1}{2}$, was tabulated in TABLE I on page 12.

However, if the null hypothesis is true, and if F(x) is continuous, the cumulative distributions of Db and Dn must be fairly close together. Due to the fact that F(x) is discrete, comparision between the two distributions shows that the value of $k/\sqrt{n}$ for Db is significantly lower than that of Dn.

Figure 2 shows the discrepancies between the two distributions at lower probability levels. In other words, if one uses the critical value of Dn with given level of significance, say α, to test for goodness of fit in binomial distribution, the actual α level is significantly lower than the original choice.

Figure 2. Comparision of Db and Dn distributions

n = 20

$k/\sqrt{n}$	$P(Db \geq k/\sqrt{n})$	$k/\sqrt{n}$	$P(Dn \geq k/\sqrt{n})$
.100	.5034	.231	.20
.150	.2632	.246	.15
.200	.1154	.264	.10
.250	.0414	.294	.05
.300	.0118	.356	.01

n = 30

$k/\sqrt{n}$	$P(Db \geq k/\sqrt{n})$	$k/\sqrt{n}$	$P(Dn \geq k/\sqrt{n})$
.067	.5846	.19	.20
.100	.3616	.20	.15
.133	.2004	.22	.10
.167	.0988	.24	.05
.200	.0428	.29	.01
.233	.0162		
.267	.0052		

2.3. Kolmogorov-Smirnov Test Statistic for the Multinomial Distribution

Consider the multinomial distribution $M(n;p_1, p_2, \dots , p_{10})$ stated in section 1.3.. If the probabilities, p_i's, are correct as specified in (1.7), one has the following cumulative distributions

for n = 1024 and n = 512 respectively;

1	11	56	176	386	638	842	968	1013	1023	1024
.5	5.5	28	88	193	319	424	484	506.5	511.5	512

Hence from (2.1), the test statistic for this sample can be expressed as;

$$Dm = \sup_{x} |Sn(x) - F(x)| \tag{2.6}$$

One may be able to obtain the probability distribution of Dm by direct computation using (1.1); but, as pointed out in section 1.3., direct computation becomes cumbersome. For those samples obtained in section 1.3., one computed Dm given in (2.6), hence it was possible to form a sample cumulative distribution of the test statistics Dm. The sample cumulative distribution of Dm is tabulated in TABLE II on page 15.

Comparision of the two distribution, Dm and Dn, shows that the value of $k/\sqrt{n}$ for Dm is lower than that for Dn. The reasons for the lower $k/\sqrt{n}$ for Dm may be either that F(x) is discrete or that Dm is a random variables whose asymptotic distribution is defined as (2.4). or both. Either stronger justification or proper modification is needed in order to apply the kst in the situation of discrete F(x), especially with small n.

COMPARISION OF THE CHI-SQUARE AND KOLMOGOROV-SMIRNOV TEST STATISTICS

For the application of the cst for goodness of fit, appropriate grouping is needed. Mann and Wald (13) have given a technique for

deciding on an optimum number of class intervals for the cst applications. Grouping observations into intervals for the kst tends to lower the value of $k\sqrt{n}$ in (2.2). Examples given in previous sections indicate that the $k\sqrt{n}$ for both Db and Dm are lower than those tabled, but X^2b and X^2m are good approximations to the chi-square distribution. Hence for the discrete distribution, the kst is conservative.

The kst is correctly used only when the distribution is continuous and completely specified. The distribution of the Dn is, therefore, not known when certain parameters of the population have to be estimated from the sample, but one may safely conclude that the discrepancy between the sample distribution and the hypothetical distribution is significant if the value of Dn exceeds the table value (15). The cst is, however, easily modified by reducing the number of degrees of freedom and can be applied to the situation where the estimation of parameters is needed.

The kst will usually require less computation than the cst. The kst treats individual observations and thus does not lose information by grouping, as the cst necessarily does. With small samples this loss of information in cst procedures is large, so use of the cst is not advisable (15).

The kst, at least 50% power level, will detect the smaller deviations in cumulative distribution than will the cst (15). In general, the power of the cst is not known (14), whereas a lower bound of power of the kst can be computed for any alternative (15). However if the kst is applied to a discrete population, nothing can be said about its power. Also the fact that one obtained lower $k\sqrt{n}$ as pointed out above, explains the reason not to use the kst in the discrete situation.

TABLE I

The entries in this table are c (1st column) $k/\sqrt{n}$ (2nd column), and one-half of the probability* (3rd column) that X^2b and Db are less than or equal to c and $k/\sqrt{n}$ respectively for each n.

n = 5

c	$k/\sqrt{n}$	p
.2000	.1000	.5000
1.8000	.3000	.1875
5.0000	.5000	.0313

n = 10

c	$k/\sqrt{n}$	p
.4000	.1000	.3770
1.6000	.2000	.1719
3.6000	.3000	.0547
6.4000	.4000	.0107
10.0000	.5000	.0010

n = 15

c	$k/\sqrt{n}$	p
.0667	.0333	.5000
.6000	.1000	.3036
1.6667	.1667	.1509
3.2667	.2333	.0592
5.4000	.3000	.0176
8.0667	.3667	.0037
11.2667	.4333	.0005
**	**	**

* Table of the Binomial Distribution
Department of Commerce, National Bureau of Standard
Applied Mathematical Series No. 6.

** The larger values were omitted for n = 15, 20, 25, 30.

TABLE I (cont.)

n = 20

c	$k/\sqrt{n}$	p
.2000	.0500	.4119
.4000	.1000	.2517
1.8000	.1500	.1316
3.2000	.2000	.0577
5.0000	.2500	.0207
7.2000	.3000	.0059
9.8000	.3500	.0013
12.8000	.4000	.0002

n = 25

c	$k/\sqrt{n}$	p
.0400	.0200	.5000
.3600	.0600	.3450
1.0000	.1000	.2122
1.9600	.1400	.1148
3.2400	.1800	.0539
4.4800	.2200	.0216
6.7600	.2600	.0073
9.0000	.3000	.0020
11.5600	.3400	.0005

TABLE I (cont.)

n = 30

c	$k/\sqrt{n}$	p
.1333	.0333	.4278
.5333	.0667	.2923
1.2000	.1000	.1808
2.1333	.1333	.1002
3.3333	.1667	.0494
4.8000	.2000	.0214
6.5000	.2333	.0081
8.5333	.2667	.0026
10.8000	.3000	.0007
13.3333	.3333	.0002

TABLE II

The entries in this table are c (1st column), $k/\sqrt{n}$ (3rd column), and $P(X^2m \geqslant c)$ and $P(Dm \geqslant k/\sqrt{n})$, (2nd and 4th column) respectively for n = 512 and n = 1024.

n = 1024

c	$P(X^2m \geqslant c)$	$k/\sqrt{n}$	$P(Dm \geqslant k/\sqrt{n})$
1.646	1.00	.0068	1.00
2.032	.99	.0087	.96
2.733	.96	.0107	.88
3.490	.92	.0136	.79
4.594	.78	.0146	.71
5.527	.70	.0156	.67
7.344	.48	.0166	.56
9.524	.26	.0175	.50
11.030	.17	.0195	.41
13.362	.06	.0214	.32
15.507	.03	.0234	.21
18.168	.01	.0263	.10
20.090	.00	.0322	.05
26.125	.00	.0425	.01
		.0509	.00

c	$P(X^2 \geqslant c)$	$k/\sqrt{n}$	$P(Dn \geqslant k/\sqrt{n})$
11.030	.20	.0334	.20
13.362	.10	.0356	.15
15.507	.05	.0381	.10
18.168	.02	.0425	.05
20.090	.01	.0509	.01

TABLE II (cont.)

$n = 512$

c	$P(\chi^2_m \geqslant c)$	$k/\sqrt{n}$	$P(D_m \geqslant k/\sqrt{n})$
1.646	.99	.0078	1.00
2.032	.98	.0117	.96
2.733	.93	.0136	.93
3.490	.88	.0175	.79
4.594	.73	.0195	.68
5.527	.59	.0214	.58
7.344	.43	.0253	.44
9.524	.30	.0312	.33
11.030	.23	.0332	.22
13.362	.06	.0390	.15
15.507	.04	.0410	.11
18.168	.01	.0429	.09
20.090	.01	.0449	.03
26.125	.00	.0546	.01
		.0601	.00
		.0720	.00

c	$P(\chi^2 \geqslant c)$	$k/\sqrt{n}$	$P(D_n \geqslant k/\sqrt{n})$
11.030	.20	.0473	.20
13.362	.10	.0504	.15
15.507	.05	.0539	.10
18.168	.02	.0601	.05
20.090	.01	.0720	.01

APPENDIX

Generation of Hypothetical Distribution by Computer

One selects two 2k-digit numbers say m and n, for k sufficiently large (say 5 or larger), multiplies m by n, and extracts the middle 2k digits, which replace n. The 2k digits extracted from the middle of the product of m by n, has the rectangular distribution R(1/2, 1) (17).

Repetition of the above process will give as many random numbers as one wants. Let's call this random number r, then generate 10 r's and compare them with 1/2. Let x_i be the number of r's which exceeds 1/2, such that $x_i = 1$ if $x = i - 1$ and $x_i = 0$ otherwise, then $(x_1, x_2, \ldots, x_{10})$ has multinomial distribution $M(1; p_1, p_2, \ldots, p_{10})$ where p_i's are given in (1.7).

REFERENCES

(1) Pearson, Karl.
"On a criterion that a system of deviations from the probable in the case of a correlated system of variable is such that it can be reasonably supposed to have arisen in random sampling" Phil. Mag., vol. 50, pp. 157-175. 1900.

(2) Wilks, Samuel S.
Mathematical Statistics. New York: John Wiley & Sons, Inc., pp. 259-262. 1961.

(3) Cramer, Harold
Mathematical Methods of Statistics. Princeton University Press. 1961.

(4) Pearson, E. S. and Hartley
Biometrica Tables for Statistician. Cambridge University Press. 2nd Ed. vol. 1, 1954.

(5) Metropolis, N. & Ulam, S.
"The Monte Carlo Method", Journal Amer. Stat. Ass., vol. 44, pp. 335-342. 1949.

(6) Kolmogorov, A.
"Confidence limits for an unknown distribution function", Annal Math. Stat., vol. 12, pp. 461-463. 1941.

(7) Massey Jr., F. J.
"A note one the estimation of a distribution function by confidence limits", Annal Math. Stat., vol. 21, pp. 116-119. 1950.

(8) Birnbaum, Z. W.
"Numerical tabulation of the distribution of Kolmogorov's statistics for finite sample size", Journal Amer. Stat. Ass., vol. 47, pp. 425-441. 1952.

(9) Birnbaum, Z. W. and Tingey, F. H.
"One sided confidence contours for probability distribution functions", Annal Math. Stat., vol. 22, pp. 592-596. 1951.

(10) Feller, W.
"On the Kolmogorov-Smirnov limit theorems for empirical distribution", Annal Math. Stat., vol. 19, pp. 177-189.1948.

(11) Smirnov, N. V.
"Table for estimating the goodness of fit of empirical distribution", Annal Math. Stat., vol. 19, pp. 279-281. 1948.

(12) Schmid, Paul.
"On the Kolmogorov and Smirnov limit theorems for discontinuous distribution functions", Annal Math. Stat., vol. 29, pp. 1011-1027. 1958.

(13) Mann, H. B. and Wald, A.
"On the choice of the number of intervals in the application of the chi-square test", Annal Math. Stat., vol. 13, pp. 306-317. 1942.

(14) Massey Jr., F. J.
"A note on the power of a non-parametric test", Annal Math. Stat., vol. 21, pp. 440-443. 1950.

(15) Massey Jr., F. J.
"The Kolmogorov-Smirnov test for goodness of fit", Journal Amer. Stat.Ass., vol. 46, pp. 68-78. 1951.

(16) Birnbaum, Z. W.
"On the power of one-sided test of fit for continuous probability function", Annal Math. Stat., vol. 24, pp. 448-489. 1954.

(17) Householder, A. S.
Principles of Numerical Analysis. New York: McGraw-Hill Book Co., Inc. 1953.

A STUDY OF CHI-SQUARE AND
KOLMOGOROV-SMIRNOV TESTS

by

CHONG JIN PARK

B.S.,B.A., University of Washington, 1961, 1962

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics and
Statistical Laboratory

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1963

ABSTRACT

The chi-square test and the Kolmogorov-Smirnov test are widely used for testing goodness of fit. The former can be applied in situations where the population has either a continuous or discrete distribution, and the latter can be correctly used only in situations where the population has a continuous distribution.

Since the Kolmogorov-Smirnov test is based on the assumption of a continuous distribution, it was of interest to see whether this test may be applied in a situation where the distribution is discrete. Two completely specified discrete distributions were considered.

The exact probability distributions of the chi-square test statistic and the Kolmogorov-Smirnov test statistic were tabulated and compared for small samples ($n \leq 30$) from a completely specified binomial population.

The comparision of the two test statistics was extended to large random samples ($n = 512, 1024$) from a completely specified multinomial population. The approximate distributions of the chi-square test statistic and the Kolmogorov-Smirnov test statistic were obtained by the Monte Carlo technique.